IASSIST
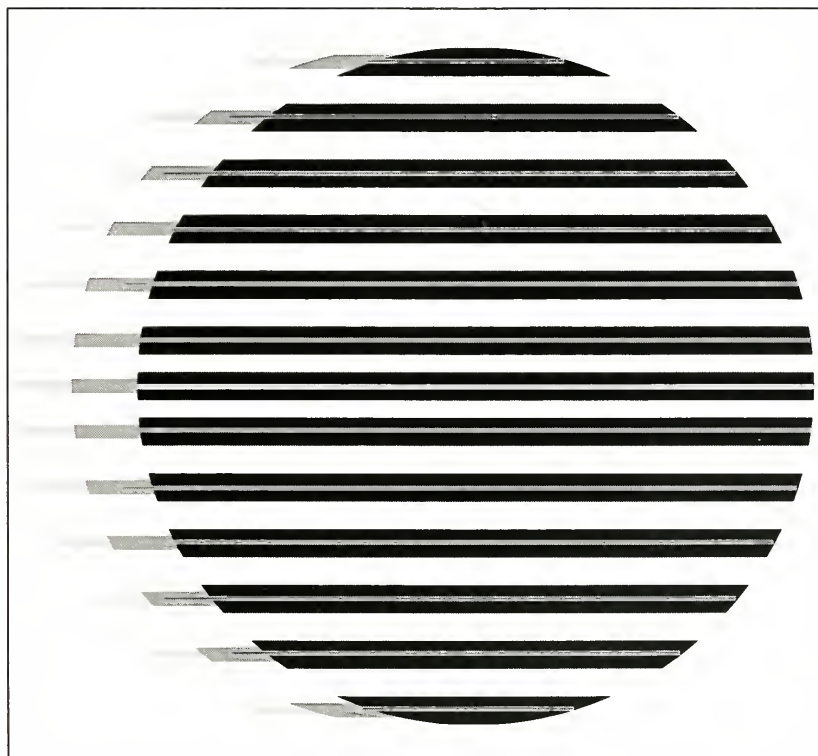
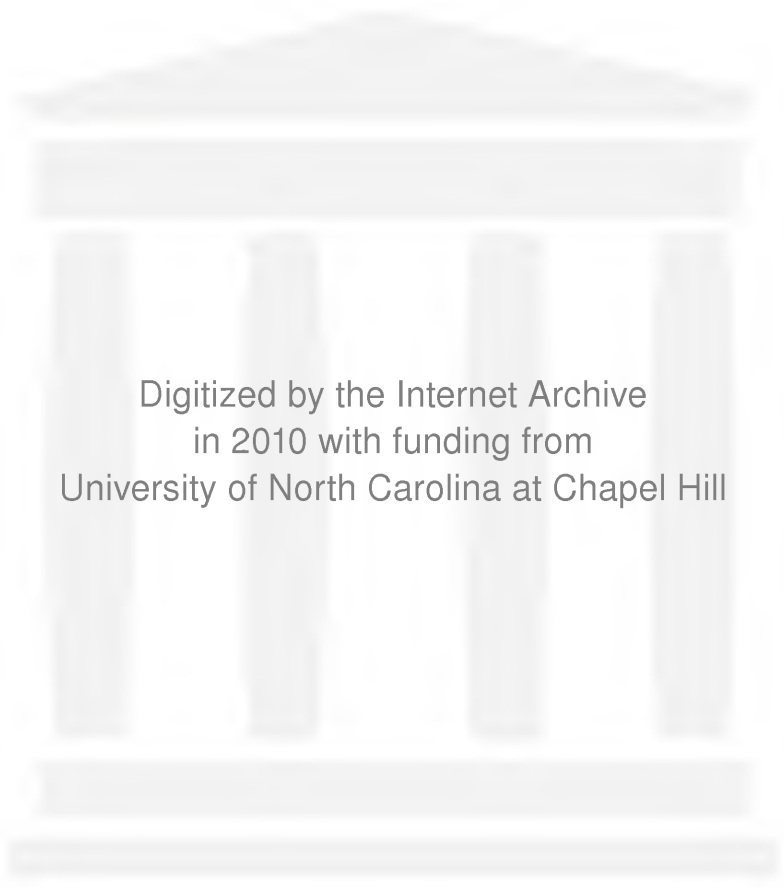## Q U A R T E R L Y

Digitized by the Internet Archive
in 2010 with funding from
University of North Carolina at Chapel Hill

http://www.archive.org/details/iassistquarterly211inte

IASSIST QUARTERLY

Information for Authors:

Editors

Production

CONTENTS

Volume 21   Number 1   Spring 1997

FEATURES

News

Data Archiving in Africa: The South African Experience

Introduction

The South African Data Archive (SADA) was established in 1993 by the Centre for Science Development (CSD) of the Human Sciences Research Council (HSRC) in Pretoria. The first staff member of SADA joined the HSRC in January 1994. As a data archive SADA's primary function is to locate, acquire, store and disseminate mainly quantitative machine-readable research data in the humanities and social sciences.

South Africa's past has divided the research community in the country into two major groups: (i) the expert minority comprising mainly advantaged institutions (some universities, technikons and research councils; and the private sector); and (ii) the disadvantaged majority comprising largely universities, technikons and other institutions in the former homelands. In the short term, the expert community will tend to play the role of "depositor" of data at SADA, while both the disadvantaged and expert communities will be users of the data.

Nineteen ninety-four (1994) was an epoch year in the history of South Africa. Apart from political renewal, it heralded a period in which researchers in South Africa are faced with immense challenges in the development of a sound science and technology system. Natural and social scientists are engaged in a number of large-scale local, national, regional and international research projects. However, the bulk of the government's funding generally goes to natural sciences research and its support facilities.

A unique challenge facing South Africa in the transformation process following the first democratic election in 1994 is to implement the National System of Innovation in Science and Technology, as defined in the recently published green paper on science and technology. The paper highlights national data gathering involving large sums of money, and suggests that the data be subjected to public scrutiny. South Africa will soon establish a National Research Foundation (NRF) which shall promote and support research through funding, human resource development, infrastructure provision and capacity building in order to facilitate the creation of knowledge, innovation and development in all fields of science and technology. The NRF draft bill promulgated earlier this

*by Maseka A. Lesaoana**

year indicates that the CSD of the HSRC will form part of the social sciences and humanities division of the NRF. However, it is not clear whether SADA and other research information facilities will also move to the NRF with the CSD. The future of SADA is shrouded in uncertainty.

Historical background

In February 1993 Per Nielsen, the then Director of the Danish Data Archive (DDA), was invited to the HSRC as a consultant to undertake a two-week feasibility study on the viability of establishing a data archive in South Africa. He wrote a detailed report on his findings, including suggestions based on the experiences of data archives in other countries. Nielsen found the HSRC a suitable location for the archive for the following two reasons: (1) the HSRC has the technical infrastructure needed to improve the quality of research and to provide training at all levels countrywide; and (2) the wealth of research data at the HSRC since its inception in 1969 can be placed at the disposal of SADA for secondary analysis.

Some of the important issues raised in Nielsen's report included the following:

- standards for proper documentation of data
- rationale for establishing a data archive in South Africa
- implementation of the organizational structure
- a variety of financial models including user payments
- recovery of costs
- staff composition
- establishment of an interim advisory body
- types of data at the HSRC that could constitute the first holdings of SADA

Also discussed in Nielsen's report were the existing documentation standards, data-processing issues, international participation and co-operation, and implementation of bilateral co-operation. Nielsen's recommendations were used as a blue-print for the establishment of SADA.

Challenges and developments

Location

SADA seems to be appropriately located at the HSRC, based on the experiences of other data archives, for carrying out its data-archiving activities. Since its inception in 1969 the HSRC has accumulated a wealth of research data, mainly from numerous projects undertaken in collaboration with researchers at various universities countrywide. This provides an opportunity for SADA to liaise with these researchers and their institutions.

At the HSRC, SADA is a division within the Research Information Directorate (RID) of the CSD. Besides housing SADA, RID also maintains a set of databases on human sciences issues through which it provides bibliographical information on current and completed research projects, research and professional organizations, researchers, and forthcoming conferences.

The CSD is committed to redressing the imbalances in research opportunities, and to empowering scholars from historically disadvantaged sectors by actively supporting their participation in the research structures, activities and decision-making processes of the broader research community. It also administers the allocation of various categories of research funding and scholarships to postgraduate students and researchers in the human sciences at South African tertiary institutions and NGOs. The Directorate: Research Capacity Building of the CSD supports the expansion of institutional research capacity by developing research skills among disadvantaged scholars.

Staff

SADA started its operations in 1994 with three staff members who joined the HSRC in January, September and November respectively. Finding suitable staff in a country where data archiving and secondary analysis are new activities is not easy. Furthermore, there are no data libraries or similar facilities in South Africa, and methodology training courses are almost non-existent. The high staff turnover rate at the HSRC has also adversely affected SADA's progress. Two of the three staff members left the organization (in February 1995 and December 1996). A fourth post of administrative assistant was filled in August 1996.

SADA staff were fortunate to be able to work with experienced data archivists. Repke de Vries of the Steinmetz Archive in the Netherlands visited SADA twice: (1) in October 1994 for a period of five months. The aim of this visit was to offer advice during the initial phase of the development of the archive; and (2) in May 1996 for a period of two weeks in order to study the developments implemented since his previous visit. Shalane Sheley of ICPSR (Inter-university Consortium for Political and Social Research) was contracted to SADA for a period of one year ending in September 1996.

In addition to the four posts originally approved, three new posts were added during 1996. Two posts were filled in January 1997 and one in March 1997, and two are vacant.

While many lessons can be learned through electronic discussions, appropriate technologies to meet the needs of the majority of researchers in a specific country can best be established through interacting directly and exchanging views with the researchers concerned.

The research community

A data archive needs to define clearly the research community for which the services are to be provided, both at supply and demand points. SADA, as a national data archive, has stratified its research community into: (a) the academic community comprising universities, technikons, and training colleges (there are 21 universities and 15 technikons in South Africa); (b) various government departments including the Central Statistical Service, health, education, police, prisons and correctional services; (c) parastatal organizations such as the Human Sciences Research Council (within which SADA is housed), the Medical Research Council, the Social Sciences Development Forum, the Development Bank of Southern Africa, and several economic and social research institutes in the country; (d) research NGOs such as the Community Agency for Social Enquiry (CASE) and the South African Labour and Development Research Unit (SALDRU); and (e) the private or commercial sector, including the mining industry and the Association of Market Research Organizations (AMRO).

By defining the research community, SADA can locate the type of studies undertaken and identify who should receive the archive's services and what types of data are available or needed. Knowing the research community also assists in the establishment of advisory committees and indicates the target groups to be consulted when assessing the delivery of services by the archive.

Meeting the needs of researchers

Data archives store data in such a way that they meet the needs of the majority of their researchers. Often suppliers (depositors) of research data are also users at the demand point. However, at the demand point new researchers also become involved. An archive's developmental stages should keep pace with researchers' interests.

Computer technology plays a major role in data archiving. New developments in computer technology occur regularly, and data archives are faced with the challenge of keeping abreast of these technological advances. The recent switch from the mainframe to the UNIX environment is a typical example. Because of the increasingly large amounts of data involved in research, new distribution and storage mediums and new statistical software packages are emerging. CDROMs and cartridges have gained recognition, and due to their large storage capacity they are

employed for storing large datasets in place of the still widely used diskettes. Mainframes' 9-track tapes are being phased out. Data are stored and disseminated in a highly compressed form, and the Internet has brought about new and user-friendly developments in computer technology. The data archivist's job is to meet users' needs effectively, while at the same time ensuring that the data in the archive are preserved for long-term usage.

SADA has devised a user survey form (also available on the Internet) to determine what tools are most often employed by researchers in their analyses. This will assist SADA in catering for its users' needs. To date the response rate has been low, perhaps because in some cases researchers are not yet familiar with computer tools.

Since 1994 SADA has organized workshops at major research centres in the country, mainly at historically disadvantaged universities (HDUs) where the lack of infrastructure has proved to be a major problem. These workshops are aimed at creating awareness of the existence of SADA and its activities.

SADA publishes a newsletter, SADA News, and a draft SADA Guide (catalogue of holdings) to inform the research community of its activities and developments in data archiving. The study descriptions of SADA's collections are also accessible on the Internet (World Wide Web).

Financial and time constraints
Financial constraints impede the establishment of any archive and the supply of data to the archive. The provision of cost-effective services by SADA is not (yet) understood by funders and owners of data in South Africa since secondary analysis itself is not well understood. Even the data that are available are often not properly documented. Data suppliers frequently request funds to enable them clean up their data.

SADA advisory committees
After the visit by Per Nielsen in 1993, an Interim Advisory Committee (IAC) of SADA was set up to offer advice on the establishment of a data archive. The IAC was replaced by the SADA Board, also an advisory body, in February 1996. The board currently consists of 17 elected members representing the research community in South Africa. The SADA Board reports to the HSRC Council. An executive committee of the board was elected in November 1996.

General progress
SADA has progressed well - thanks largely to the help of the established data archives of Europe and the USA. For instance, the above-mentioned visit by the then Director of the DDA, Per Nielsen, contributed greatly to SADA's understanding of data archiving. His investigations into research studies already undertaken in South Africa, guided

SADA's acquisition of potential holdings. He also assisted in registering SADA as a member of the International Federation of Data Organizations (IFDO), which led to a visit by the head of SADA to a number of IFDO data archives: the ICPSR in the US, the ESRC Data Archive, the DDA, the Swedish Data Archive (SSD), and the Steinmetz Archive (STAR). The purpose of these visits was to learn how other data archives operate in order to design a suitable model for SADA. It was during these visits that (i) SADA became a member of the ICPSR, and (ii) negotiations took place with Repke de Vries of STAR, resulting in his visiting SADA for a few months to advise the staff at the take-off phase. He gave valuable tips on all aspects of data archiving and data management.

During the planning stage, a CDROM drive, SAS & SPSS, DBMSCOPY and Folio Views were acquired, and the Internet facilities were set in place.

Conclusion
The earliest data archives were established in Europe and North America in the 1960s. Through international associations such as IFDO, IASSIST and CESSDA these archives established guidelines for data archiving and shared their knowledge on the collection, storage, documentation and dissemination of data. Newly established data archives are fortunate to be able to share this knowledge. However, for any data archive general considerations as well as unique considerations pertaining to the specific country have to be taken into account.

General considerations include:

> **Location:** in terms of wealth of and access to research data; institutional credibility; and technical and infrastructural support;

> **Staff:** qualified staff to run the activities of the archive. Initially fewer staff with broad knowledge base are needed;

> **Networking:** networking with researchers and research institutions that produce data is vital. Advisory committees formed with representatives from research institutions can help strengthen networking relationships;

> **The science of data archiving:** through international links such as IFDO, a new data archive can learn the science of data archiving, such as the procedures for acquisition, processing, storage and dissemination of data; and

> **Funding:** this is a major problem. Data archives are not profit-making facilities, and often depend on government funding.

Most specific lessons learned by SADA stem from the economic and political changes currently taking place in South Africa. Data archiving activities are directly linked to the research capacity building activities at higher institutions of learning. Due to apartheid, the education system in the country has been badly skewed. A few privileged institutions had the capacity to undertake research while the majority did not have the resources to do so. Sharing of research between these two groups did not take place. SADA is consequently faced with the challenge of rectifying this situation and also of promoting the new disciplines of data archiving and secondary analysis throughout South Africa and Africa in general.

Other issues to be considered by new data archives:

Research practices: these differ from country to country. Some researchers do not want to give other researchers access to their data for reasons ranging from not being accustomed to the culture of sharing, to wanting to make money by selling their data;

Poor research training programmes: unlike in western countries, South Africa has poor methodology programmes. Summer schools in Northern America and Europe, for example, are held to strengthen the research capacity of data archives and data libraries;

Collaborative research in the region: joint activities by South African researchers are only now starting to take place. Communication breakdown between African countries makes collaborative research difficult. In Europe, for example, the Eurobarometer studies and International Social Survey Programmes are well-known research studies undertaken jointly by European countries.

While some good lessons can be learned from visiting established data archives abroad, the best option is to invite experienced data archivists to the new archive at home. Technology at developed data archives is already at an advanced level and it may be difficult, if not impossible, to implement new practices at the new archive where the infrastructure may be poor or not compatible.

IASSIST and IFDO conferences have traditionally been hosted by Canada, Europe and USA in that order. The mission, objectives and activities of these bodies should be reviewed. One objective should be to draw in more participants, particularly from the third world countries and countries not previously included. Efforts should be made to host such conferences on other continents and to provide funding for participation by researchers in third world countries, thereby increasing membership and building research capacity around the globe.

Statistics Norway and the Social Sciences

*by Johan-Kristian Tønder**

Objective
Statistics Norway (SSB) shall prepare and distribute statistical information on the Norwegian society. The official statistics shall co-operate with research and analysis in order to monitor and analyse economic and social conditions as well as resource and environmental conditions. It shall thereby provide the public, the industry and trade sectors, and the authorities, with information on the society's structure, developments and mode of operation.

Statistics and complementary analysis from Statistics Norway is a common property which should be used by as many people as possible. Statistics and analysis shall therefore be easily available to all, that is, to individuals, authorities, political and other organisations, educational institutions, the media, establishments, etc.

The Institution
Statistics Norway is a professionally autonomous institution. This means that Statistics Norway has full responsibility for the professional contents of the statistics and analysis. Statistics Norway is completely independent in deciding what official statistics shall be published and when and how this should be done. This independence from the authorities and interest groups is crucial as confidence and authority are essential prerequisites for official statistics and also for Statistics Norway to fulfil her role in the Norwegian and international society.

The Situation
During the last decade, Statistics Norway has had a real cut in her basic allocations from the national budget. The number of employment positions allowed by the national budget have gone down from over 700 to just under 600. In principle, the area of statistics covered by government funding should include a core of statistics and research subjects, defined from what is considered to be most relevant for the Norwegian society. That goal is difficult to fulfil with decreasing funding.

In certain cases, the establishment of new areas of statistics or the reinforcement and improvement of existing areas of statistics are not covered by current government funds. This will require binding long-term co-operation agreements with the ministries that have a special need for improvements in statistics and research for monitoring and for policy-making in their areas of responsibility.

These ministries must cover costs incurred during the development and operating phase of this increased production of statistics.

Pricing policy
The pricing policy of SSB as regards distribution and commissioned assignments shall be consistent with the ruling principle that statistics and analysis should be commodities that are easily available to various users.

For commissioned assignments, which may consist of the development of new statistics, preparation of statistics or research and analysis, the guidelines require that the client covers SSB's real costs, meaning full cost absorption.

In certain cases, users may require more detailed or processed figures outside the scope of statistics produced by government funding. The extra work required to provide such information shall be priced at its marginal cost when the statistics is distributed.

Data collection
At Statistics Norway is done by:

- sending questionnaires to "all" persons and establishments
- sending questionnaires to or interviewing a selection of persons or establishments
- setting up at SSB, registers of persons and establishments
- gathering information on persons and establishments collected by other public bodies.

Organisation of data
The objective of the organisation of data is that all users shall have simple and good access to official statistics of Norway. One way of answering the challenges from the major users would be to distribute data via databases where the user can make his own choice of pre-defined connections. The databases shall be user-friendly and shall contain metadata (descriptions and definitions of data etc.) while safeguarding the protection of privacy. In certain

areas, such databases are under development.

Data protection
In the production and distribution of statistics, confidentiality and the protection of privacy are strongly emphasised. Safeguarding the anonymity of the client is a vital prerequisite for the activities of Statistics Norway. Individual data on persons or establishments/enterprises should be treated confidentially and should only be used for statistical purposes and in accordance with the regulations laid down by the Norwegian Data Inspectorate.

Research and statistics
Statistics Norway carries out research on her own data. Research activities shall cover research for social planning in the areas where Statistics Norway has a particularly central statistical responsibility. As such, we compete with others who use statistics in their analysis activities. However, the research is important and necessary for further development and improvement of the production of statistics in the various areas.

Distribution to researchers
Social researchers are major users of our publications and databases. In many cases, researchers provide feedback that contributes towards quality control and improvements of the statistics. However, research-projects often requires other statistics than what we present in standardised form. In such cases, researchers may order separate data selections to suit their purpose. Sometimes, a researcher may like to combine his own data with our data. In such cases SSB combines the data and produces statistics or anonymised data files for the researcher.

However, research often requires other types of data than normally produced in statistics. In particular, this concerns event history data where one follows the same individual over time and in different social situations such as education, employment, social security benefits, etc. As government funds allocated for the production of statistics are decreasing, it is difficult to get funding for event history databases within this form of finance. We therefore have to try to find other sources of finance. Among others, the research sector itself.

The NSD Agreement
In 1993 Statistics Norway and the Norwegian Social Science Data Services (NSD) entered into an agreement, with the purpose of providing researchers and students with the easiest possible access to as much of SSB's data as possible without violating the rules laid down by Statistics Norway and without reducing the degree of protection of sensitive data.

The agreement covers two types of data. Firstly, statistical data tables ready to be published from SSB's statistical databases or specially prepared upon NSD's request.

Secondly, anonymised individual data from SSB's sample surveys or other data from persons or establishments.

For these services the agreed price is lower than the marginal cost that is usual for SSB's services. I must confess that due to inadequate capacity, SSB cannot always fulfil the agreement as quickly as desired.

The agreement says that SSB is committed to keep NSD informed about the surveys and statistics that are produced and to inform students and researchers about NSD's services. NSD shall in return inform them about the services that SSB can offer and ensure that SSB is cited as the source of data published in NSD's own publications or presented in reports, periodicals etc, by researchers .

The agreement has worked positively for both NSD and SSB. NSD has a reliable and reasonable access to important data for social research. SSB has been relieved of part of its services for students and researchers and SSB's data and the possibilities lying in these data have become familiar for students and researchers.

However, in SSB's view, the arrangement has two disadvantages. Firstly, when students are trained to use NSD, the result is that many of them do not approach SSB when they enter occupational life. They still want to use NSD. Consequently, NSD's subsidised service which is arranged for students and researchers, is demanded by groups for which it was not intended. This creates uncertainty about the division of roles between NSD and SSB.

Secondly, many of NSD's users are not always giving reference to SSB as the source of their data. Often, reference is only made to NSD. Subsequently, there is less appreciation for how important a well-established statistical agency is for research and education as well as for the society. The official statistics are taken for granted. At a time when there is a battle for public resources, the lack of understanding may have the consequence that statistics looses priority. Here, I refer to basic official funding of Statistics Norway as I mentioned before.

However, altogether, the conclusion is that the agreement works positively for both partners.

A non-authorised version of the agreement is enclosed this paper.

Data preparation at SSB for researchers
As I mentioned many users need specially prepared tables for their purpose and these tables are not always available from NSD's service. These users therefore have to contact SSB for special tables prepared. Very often the researchers would like several data sources to be combined in order to produce tables. In such cases, the researcher has to deal

with several divisions and in order to prevent this, SSB shall establish a fixed contact point for such services. This contact point shall be set up at our division for Population and Housing Census 2000 as it has been decided that the Norwegian population census shall be carried out by combining various registers. Subsequently, this division has to build up the competency required for combining registers and this may also be of use to research. We suppose that in many cases NSD will be the institution that demand such data on behalf of many researchers and students.

Perspectives
SSB has a three-pronged strategy for the provision of services for social research.

The first is (of course) "more and better statistics" especially in fields of statistics that can fill out the uncovered areas we have today, both within national accounting and within statistics on demography and social welfare.

Secondly, to continue to direct our efforts towards co-operation with NSD in the distribution of statistics to students and researchers.

We hope that this co-operation can be further developed in connection with the third prong of the strategy which is to increase the efficiency of our readiness to combine registers for research purposes. In this context, we also have to look into the development of databases for event history on persons. These databases should be built on many statistical sources and more annual volumes of these sources. Here, however, SSB, social research and social planning probably have to come together to find an appropriate form of finance for what could become the «new gold mine» of Norwegian social research.

Non-authorised version

AGREEMENT
between
Statistics Norway (SSB)
and the
Norwegian Social Science Data Services (NSD)
on
THE DISTRIBUTION OF DATA FROM STATISTICS NORWAY

1. Purpose
a) The purpose of this agreement is to give researchers and students the easiest possible access to the largest possible amount of SSB's data without violating the current regulations and without relaxing the protection of individual data.

The term researchers refers specifically to researchers affiliated with a university/ college/ regional college in Norway or abroad and the institutional sector in Norway. The social science institutional sector is to follow the current definitions established by the Norwegian Research Council. Enclosure 1 comprises all those on NORAS' list at the turn of the year 1991. Researchers from similar institutions in other sectors shall have the same rights. If there is doubt about whether an institution belongs to the institutional sector, the matter shall be resolved by the Norwegian Research Council. All persons enrolled at a university/ college in Norway or abroad are considered as students.

b) The distribution of statistics to other public and private services is outside the scope of this agreement. However, NSD may distribute data on condition that NSD pays an additional fee to SSB (enclosure 3). Individual data shall not be distributed to persons or institutions outside the primary target group without SSB's special consent.

c) In addition, SSB and NSD may prepare separate co-operation agreements for the delivery of integrated program and data packages intended for various user groups.

2. Scope
This agreement covers the delivery of the following types of data:

Group I. Statistical tables ready to be published.

Group II. Individual data, rendered anonymous, from SSB's sample surveys or other individual data (persons, establishments) also made anonymous.

3. Delivery procedures
Group I data:
If data can be delivered from SSB's statistical databases, they are produced as standardised tables as soon as the underlying figures are registered in the database, and at an agreed price. Enclosure 2 gives a summary of prices and the types of data to be regularly delivered in this way.

NSD may choose not to have data delivered regularly, but may place an order in each instance. This is specified in enclosure 2. Alternatively, NSD may be granted on-line access to retrieve data themselves from SSB's databases.

In the case where NSD orders a set of data that does not exist in the databases but requires special processing, an estimate of the additional work required to deliver the data shall be calculated. NSD shall confirm the order before the assignment is carried out.

Group II data
NSD shall apply to SSB for the transfer of data from individual surveys. SSB shall evaluate the application and give approval, subject to SSB's concession from the Data Inspectorate. Alternatively, SSB may propose changes in the selected data and prepare an estimate for the additional work required to deliver the data. Researchers shall submit an application to NSD for access to the anonymous data. The application shall specify the project, etc. where the data shall be used. NSD shall process the application and approve or reject it. Before access is given, the user must sign a special pledge of confidentiality.

NSD shall keep a record of accepted applications. Furthermore, this record shall contain project information and a list of any publications that may be based on the material. NSD shall send an update of these records to SSB every six months.

For anonymous individual data that is not transferred to NSD, or tables with confidential information, NSD must apply for transfer in each instance and SSB shall determine the conditions that apply for the set of data in question.

4. SSB's rights and obligations
Restrictions on distribution and use of any particular set of data and possible claims of return, shall be determined by SSB at any time.

SSB shall keep NSD informed on all sets of data from sample surveys. As soon as the questions are finally agreed upon for each sample survey, SSB shall immediately

dispatch a questionnaire and give a probable delivery date for the data. More detailed documentation on the data shall be delivered on NSD's request, as far as SSB finds this possible in practice.

SSB shall provide regular updates on contents and changes in SSB's databases.

SSB may demand that reports and documentation on SSB's data, that are prepared and distributed by NSD, be submitted to SSB for approval.

SSB may request copies of delivered data and complementary documentation, against reimbursement.

SSB shall contribute to inform prospective user groups about NSD's services.

SSB is committed to deliver correct data that is released for delivery as well as documentation required to use and interpret the data. If any errors are detected during the transfer of data, SSB shall immediately arrange for correct data to be transferred.

SSB may make data originally prepared for NSD, available also to other users without notice to NSD.

SSB reserves the right to verify that the data is used in accordance with the conditions and requirements stipulated in this agreement.

5. NSD's rights and obligations
NSD has the right to distribute data to the user groups specified under point 1a.

NSD may use data from SSB in their own analysis and assignments for users specified under point 1a.

For distribution of tables to other users [administration, private users etc. (1b)] NSD shall pay a separate fee to SSB for the data used, in accordance with enclosure 3.

When integrated as a part of teaching packages/ test data connected to NSD-STAT, NSD may also distribute data from SSB to users other than those specified under point 1a. For distribution to non-primary users, reimbursement for the data shall be made in accordance with enclosure 3, or possibly in accordance with a separate agreement.

For all use of SSB's data, SSB shall be clearly cited as the source, both as a general reference and in association with individual tables where the data is used. This condition applies both for NSD's own use and usage of data distributed through NSD.

NSD shall establish protection for the data and prepare instructions for processing and data storage. Data

protection procedures and instructions shall be approved by SSB.

NSD shall report to SSB any errors and deficiencies that are detected in the data or documentation.

NSD shall contribute to distribute information to prospective user groups about SSB's services.

NSD shall give SSB free access to all teaching packages that are prepared with data from SSB.

NSD may have on-line access to SSB-DATA by paying the same subscription fee paid as schools/ libraries.

Every six months, NSD shall send a report on usage in accordance with enclosure 4, which will serve as a basis for settling the fee for distribution to non-primary group users.

6. Contact persons
Each of the partners shall designate a contact person. Changes are to be continuously reported.

7. Validity
This agreement is valid from the date of signature and can be terminated with three months notice. Enclosures 2 and 3 may be adjusted annually.

* Paper presented at the IASSIST/IFDO 1997 Conference, May 6th May 9th Odense, Denmark.

Johan-Kristian Tønder, Assistant Director General, Department of Social Statistics

A Presentation of the ERAS Project: The Danish Unit for Registration and Storage of Medical Research Data

*by Peter Heine Jorgensen**

The new Unit was established on november the 1st 1996. Years of work and planning by the Danish Data Archives finally led to its formation half a year ago. Sadly the initiator, the director of the Danish Data Archives, Per Nielsen, died less than two months after it's formation.

The purpose of the Unit is to strengthen the registration and storage of medical research data in Denmark. The new Unit will provide a professional storage function to the medical research community and promote access to collections of medical research data for secondary analysis. The Unit is established as a 5-year collaboration project between the Danish National Research Foundation and the Danish Data Archives.

In an effort to strenghten the research development capabilities in Denmark, the Danish National Research Foundation was established in 1991. Since its formation the foundation has worked to improve the conditions for research development in Denmark in the entire scientific landscape. This is achieved by giving large and concentrated grants to unique Danish research at the international level. Several new projects have thus been started, the establishment of the Danish Unit for Registration and Storage of Medical Research data being one of the most recent. Denmark has a unique tradition for keeping and maintaining population based registers, which places the new Unit in an almost ideal environment for developing a professional archive for medical research data.

The Unit is physically located at the Danish Data Archives, as a fully integrated part of the institution. The aquisition, processing and subsequent storing of medical research data closely follows the principles already developed at the Danish Data Archives. A simple copying of the existing routines is not enough though, as the, at times, different nature and sheer volume of medical research data dictates the development of new and refinement of existing routines.

Since its establishment in 1973, the Danish Data Archives have collected data and documenta tion from social science and historic studies, and until now the institution has concentrated it's efforts on these areas of research.

Consequently medical research studies constitutes only a small fraction, about 6% of the total contents, primarily in the form of social medicine and occupational health studies.

About 8000 articles are published every year by Danish medical researchers. Although not all of these articles represent a seperate research study, even a conservative estimate of 2 articles per study gives a yearly volume of 4000 studies. Add to this a considerable backlog of old studies and you begin to realize the magnitude of the task. This necessitates the development of new, timesaving procedures for registration and storage of data and documentation. An important consideration in the development work is to ensure that the high quality of registra tion and storage of data and documentation already attained at the Danish Data Archives is preserved.

The staff of the Unit at present consists of 3 persons:

The project leader, medical doctor and Ph.D. Kirsten Kyvik, who also is a member of the management group for the Danish Twin Register, senior researcher, medical doctor and specialist in community medicine and occupational health, Peter Heine Jorgensen and informatics assistant, Birgit Wich.

It is the first time that medical staff has been employed at the Danish Data Archives and indeed at the Danish State Archives. The reason of course being inherent in the nature of the project. The logic being that a medical staff is best suited to deal with medical research data and documentation. Furthermore it is expected that this will counteract any reluctance or mistrust on the part of the donor of the medical research data. The employment of medical doctors in the staff also solves the confidentiality issue often associated with handling medical research data.

Another task presents itself to the Unit regarding confidentiality. The data may contain sensitive, identifiable information about the persons in the study base.

The Danish Data Protection Agency under the Danish Justice Department controls and regulates the legal aspects of data storage. The Agency has granted the Unit permisson

to store sensitive data from medical research. Furthermore permission to release the original data and documentation to the donors is given. Thus the primary researchers are able to continue their study on the same study base, even years after completion of the first study. Naturally other researchers will only have access to the data in an anonymous form. The Unit is at present engaged in an effort to persuade the Data Protection Agency to accept that storage at the Unit can be regarded as being equal to deletion. It is hoped that this can be stipulated in the standard agreement made between the Agency and the medical researchers.

The target of the Unit is medical research in Denmark and existing as well as coming Danish medical researchers. A broad acceptance and support from the medical research community is instrumental in achieving a succesful result, ie. the creation of professional archives for medical research data. Consequently the initial action will focus on information and dialog. Agreement regarding submission of research data to the new unit will be investigated in accordance with the Danish medical research community,

The scope of the Danish Unit for Registration and Storage of Medical Research Data is to assist researchers and research institutions in making data and documentation from studies readily available. The new Unit will perform it's own research and development with the purpose of making the new archive as well functioning as possible, thus facilitating secondary analysis. The possibility of new crosslinks between the existing social science and historic data and the new medical research data represents a unique opportunity to perform secondary analysis bridging several major research fields.

The Units field of activitivity is at present confined to Danish medical research, but the findings could have an international bearing as the principles and procedures of the fully established Unit could be implemented in other countries. A more concerted implementation could be possible in the framework of the European Union.

* Paper presented at IASSIST/IFDO '1997 Conference, May 6th May 9th Odense, Denmark.

Peter Heine Jorgensen, M.D., Specialist in Community Medicine and Occupational Health.

Developing Internet Based Teaching and Learning Resources for Social Sciences: A CSCW approach.

*by Vernon J. Gayle**

Introduction

The Internet has awesome potentiality as a global network of information and the world wide web provides a potentially efficient and effective protocol for the delivery of information. The content of the resources made available are often of particular interest to those in university settings engaged in teaching social sciences. Technologies will most probably continue to be innovative and technical advances will increase access to these resources. Whilst this will be welcomed progress it also raises a series of issues and problems especially with regards to using Internet based resources in social science university curricula.

In this paper I will argue that whilst the use of Internet resources is potentially good for university teaching and learning, the development of these resources has often been rather haphazard. The driving forces behind the development of these resources have been either from enthusiasts or technical specialists. My argument is that the Internet is too precious and important a resource to be allowed to develop in such an unmanaged fashion. I will be advancing an argument that calls for the incorporation of non–technical knowledge in development of Internet based resources that are appropriate for use in university social science teaching and learning environments.

The ideas that I intend to express are to a large extent exploratory and intended to be evocative rather than the last word on the subject. My arguments are initially premised upon the conception that the Internet can be considered simply as a piece of information technology. The Internet is a global network of information which provides a potentially efficient and effective protocol for the delivery of information. In this regard the Internet is special but if we consider the popular image of the Internet as a 'super highway' then it is possible to think of it as a giant communications technology. Whilst the Internet has some specific features I wish to contend that if it is simply considered as a communications technology the issues that pervade its use in social science teaching also pervade the use of other information technologies in this area. Given this, more general arguments about information technology and social science teaching also apply to Internet based resources.

The incorporation of information technology into teaching settings generally proceeds from a 'technology is obviously a good thing' approach. In some quarters this is considered as axiomatic. Whilst I am sympathetic to the idea of incorporating technology into higher education teaching and learning it is the abandoning of this axiom that is essential if information technology is to be successful. The motivation behind information technology in the university setting has been from, enthusiasts on the one hand and technical specialists on the other. I will refer to this as the 'technologist perspective'.

My argument is that the 'technology is obviously a good thing' approach is a flawed departure point. Due to the hegemonic domination of technical expertise held within the technologist perspective the design and implementation of new information technologies in teaching and learning environments will have very limited success. I will be advancing an argument against the 'technologist' approach that calls for the incorporation of non-technical knowledge in technological developments in university teaching and learning. This will draw upon some of the advances that have come out of Computer Supported Cooperative Work (CSCW) approaches. I will also argue that a particular theoretical conception of 'work' and of the role of social science in technological design and implementation is appropriate.

Technology and Teaching Social Science

Part of my disquiet with the technologist perspective's approach to the design and implementation of new information technology in teaching and learning environments is the generality of the argument that technology is necessarily a good thing. The lack of specificity in this approach engenders a poor understanding of the particularities of university teaching and learning. In much the same way it would be easy to talk quite generally about technology and university teaching and learning, but to avoid this pitfall I will confine this discussion to the social sciences in particular.

In the social sciences, students' experiences of information technology will mostly be in the form of desk top computing of the PC or Macintosh variety. Their introduction to computing will often form part of research

methods or study skills. The role of the computer in this instance is often not to deliver computer based or assisted learning but rather software and hardware are used as tools to undertake tasks rather than as learning technologies.

In the British context, despite funding being directed towards computer based or assisted learning, in the social sciences there has been an absence of completed and useable bespoke software packages. The failure of these endeavours is evident insofar as there are few examples that are routinely incorporated into social science curricula. These bespoke software packages are not in widespread use in departments in British higher education!

The alternative to bespoke software is the re-use of existing technology. In these endeavours software, and to a much lesser extent hardware, are directed toward a teaching and learning requirement in an attempt to add value to a specific learning experience. From my own experience and that of colleagues, these attempts to re-use existing technology are at best problematic. The re-used existing technology, is generally attempted on an ad hoc basis and its development is time consuming, and labour intensive. In the majority of cases the development has not proceeded from a clear pedagogical requirement and the end products lack the sophistication required to deliver the high quality educational experience that is the hall mark of universities. The 'value added' nature of these endeavours is not necessarily tractable, especially when cost is entered into the equation.

Despite the huge potentiality of technology for social science teaching and learning, attempts to introduce bespoke software and to re-use existing technology are ill conceived. This is due to the taken for granted assumptions about the actual nature of the teaching and learning environment and the lack of a comprehensive empirical understanding of the processes that are in motion. This in turn leads to a shallow understanding of the consequentiality of the introduction of new computer based technologies in social science teaching.

It is highly likely that a similar unsatisfactory situation could arise with Internet based resources. The Internet offers social science teachers and students a global network of information resources. Internet based resources are potentially good for university teaching and learning but the development of these resources has often been rather haphazard. Once again the driving forces behind the development of these resources have been either from enthusiasts or technical specialists. The Internet is too precious and important a resource to be allowed to develop in such an unmanaged fashion.

In the world of commerce and industry many technical endeavours which are based around personal computing have not furnished adequate results. This has been due the

technology failing to pay sensitive account to what I shall term as the 'innate sociality' of the environments into which they are being introduced. This parallels the situation in social science teaching in universities. One solution to the problem of new technology and the workplace has been the development of Computer Supported Cooperative Work (CSCW) as a design paradigm. I do not wish to argue that the work environment in the world of commerce and industry is the same as it is in higher education, although there are some obvious similarities at a generic level. I maintain that a particular theoretical sociological conception of 'work' can inform CSCW design and this is appropriate to the design and implementation of computer based learning technologies in general and this extends to the development of Internet based resources for social science teaching.

In the next part of the paper I will introduce the idea of Computer Supported Cooperative Work (CSCW). The material relates to the development of technology more generally and is not restricted to either teaching and learning or the internet. This will provide a context within which my position can later be developed.

The Problem of Human Computer Interaction
Fundamental to an understanding of the propriety of CSCW is an appreciation of the problems of a Human Computer Interaction (HCI) approach. HCI was a new and radical approach to systems design that achieved prominence in the 1980's and it sought to provide a better cognitive coupling between human users and computers (Bannon 1989). CSCW can reasonably be considered as a response to the failings of Human Computer Interaction (HCI) approaches.

HCI is a general framework for innovation aimed at developing interaction techniques, analysis methods, software and computer systems within a controlled context in order to create enhanced products. HCI endeavoured to go beyond simply providing improved surface characteristics, and hoped to address wider issues surrounding human interaction with computers. In this sense HCI is a design and engineering science as it aims to produce artefacts of hardware and software within satisfactory frameworks of compromise that take functionality, performance and cost into account (Brooks 1990).

The HCI perspective recognised that there were human consequences to the introduction of new technology, and that how technology was developed and applied could profoundly affect work. With regard to the introduction of new technology questions of usability, applicability and acceptability were being raised. From within the HCI camp these issues were viewed as being of legitimate concern. What was considered necessary was an applied psychological dimension located within a problem centred approach which would enable HCI practitioners to

undertake research that would inform future designs[1] (Blacker & Osborne 1987).

Despite HCI being a radical new initiative the organisation of work is in fact endlessly richer and more complex than the majority of formal psychological models could have conveyed. Due to the rigid frameworks that such systems imposed, human actors were not furnished with sufficient flexibility to make the system function (Bannon 1989). Another draw back which stems from the psychological foundation is that HCI fashioned itself as a general paradigm for innovation and design in limited and controlled environments (Brooks 1990). Much of the early HCI work was confined to rather small scale controlled experiments with the presumption that the findings could be generalised to other settings (Barnard and Grudin 1988). The hoped for contribution of HCI to the design of computer systems and novel interfaces did not materialised in the 1980's (Carroll 1987). Gray and Atwood (1988) note the lack of examples of developed HCI systems. This is largely due to the inherent deficiencies in HCI approaches and what is required is an alternative theoretical and methodological orientation (Luff & Heath 1990).

Computer Supported Cooperative Work
The expression CSCW is a comparatively new one in the information technology vocabulary and was first coined in the mid 1980's by researchers in the USA. The term was most notably used by Greif (1988) and has been applied as an umbrella term which takes in anything to do with computer support for activities involving more than one person. An alternative terminology to CSCW includes the expressions 'groupware' and 'workgroup computing' (Clark and O'Donnell 1991).

CSCW is very much a generic term. Bannon (1990) argues that despite disagreements about specific detail most CSCW practitioners would agree with Lyytinen (1989) who asserts that CSCW is an attempt to place emphasis on both the distinctive qualities of cooperative work processes and on questions of systems design.

CSCW takes from its point of departure the visibly processual character of social activity (Harper & Randall 1992). The organisation of situated action is an emergent property of the interactions between actors and their environments (Suchman 1987). Settings engender a specificity unique to their social organisation. Sociological inquiry within CSCW must not be ad hoc or abstract and divorced from examination of the specificity of the setting. We must attempt to study settings and explore the coordinated tasks that computers might support in the context and settings which they occur (Luff and Heath 1990). It is fundamental to examine the natural settings where tasks and activities exhibit their sociality (Bannon & Schmidt 1991).

The role of an effective CSCW system as its name suggests is to support the cooperative nature of work. Hirscheim and Klien (1989) assert that the good system must not be designed in what they term as the 'usual sense', but has to be designed and developed within the framework of the social interactions that are embedded in the environment in which the technology is to be incorporated. The caveat that must be issued here is that in no sense is there an objective set of criteria that form a typology for an effective system. The system must be developed within what they term as the 'user's perspective'. A CSCW system is not however simply an electronic cloning or duplication of a working environment. In contrast it is a pragmatic attempt to support the cooperative tasks of work in context within its natural social and physical environment.

In terms of sociological inquiry, ethnography is the tool of sociological research most applicable in CSCW endeavours. As with all research methods ethnography has advantages and disadvantages but the potency, in the CSCW context, is that it depicts the activities of social actors from their own perspective. This challenges the preconceptions that alternative social science approaches often bring to phenomena (Hammersley and Atkinson 1983). Ethnography is not simply description, rather, it is detailed explication. It is about capturing the real movement of experience in the concrete world. Ethnography achieves something which theory and commentary in the majority of cases cannot - namely it presents human experience without minimising it and without making it a passive reflex of structures, organisations and social conditions. 'Human productions are all of a piece, indivisible and always summed. The metal cannot be simply smelted out from the ore of experience in human affairs' (Willis 1978 p. 180).

The use of ethnography is an attempt to ground the understanding of action in empirical evidence. An empirically based social interaction perspective is inductive from particular naturally occurring activities. Ultimately, this will produce descriptions that are accountable to evidence. Situation is crucial to the interpretation of actions but although this is fundamental and to some extent obvious, its importance could easily be overlooked.

Sociologists have long used ethnographic techniques to study work in general.[2] What is required in CSCW is an ethnographic analysis of settings that are due to be 'technologized', by which I mean where new forms of information technology are to be implemented. In the case under consideration the settings are where social science teaching and learning takes place. Straightforward ethnographies, such as those developed in the sociology of work, are not sufficient however. What is required is what Button and Dourish (1996) term as a 'technomethodologically' informed approach. This I believe will lead to the successful design and

implementation of new information technologies in social science teaching settings.

The technomethodological approach departs from the desire to make conspicuous what actors are doing when they organise the activities that they do in particular settings. It draws heavily upon ethnomethodology which turns away from the structures and theorising of traditional sociology, and concentrates instead upon the details of the practices through which action and interaction are accomplished (Button and Dourish 1996). Ethnomethodology poses the question 'what the devil does the native think they are up to?' (Anderson, Hughes and Sharrock 1989). In the case of CSCW the native is the individual in the setting about to be technologized. Examining settings by technomethodologically informed ethnography is fundamental to the exercise of CSCW if it is to be applied successfully to teaching and learning settings.

The technomethodological approach is underwritten by the work of Garfinkel[3] who is concerned with that most pervasive sociological question; 'how is it that actions recur and reproduce themselves?' He insists that this orderliness be viewed as arising from within activities themselves and the work being done by the parties to that activity. Garfinkel eschews the traditional sociological strategy of seeking to explain this orderliness and the organisation of social activities by attempting to identify causes and conditions out with the activities themselves (Benson & Hughes 1983).

Germane to this are Garfinkel's suggestions that it is evident from the availability of empirical specifics that there exists a locally produced order of work's things that make up the enormous domain of organizational phenomena.[4] He argues that the classical sociological studies of work, without remedy or alternative, depend upon these phenomena, make use of the domain and ignore it. That the domain is ignored is a systematic feature of the locally produced orderliness of work settings. Therefore the reported phenomena are only inspectably the case, therefore they are unavailable to the art of designing and interpreting definitions, metaphors, models, constructions, types or ideals and cannot be recovered by attempts, no matter how thoughtful to specify an examinable practice by detailing a generality.

The concept of the 'egological organisation' is advanced by Anderson, Hughes and Sharrock (1989) and is an ethnomethodologically informed view of organisations that begins with a bottom-up understanding of them[5]. The conception of the egological organisation departs from an inquiry into the daily or routinised experiences of individuals. The value of this approach is that it places the actor's point of view at the centre of the analysis.

By employing the concept of the egological organisation,

they develop the idea of the 'working division of labour'. In *Working for Profit* they argue that it ought to come as no surprise that actors in work settings see themselves as part of an elaborate working division of labour. From the way that they talk about their work, both to each other and to outsiders it is clear that the notion of a working division of labour is one which they use to interrelate and explicate the things that they see going on about them, on a daily basis and on ordinary occasions. These accounts depict a body of activities marshalled by 'a working division of labour'.

Technomethodology requires a sociological analysis of the organisation of social action and interaction and the organisation of work and work settings. The fullest possible description that captures the essence of the 'working division of labour' must be furnished. The thrust of technomethodology is the conception that sociological descriptions of the ways in which people routinely organize their actions and interactions can be furnished and compared to what is or is not possible using technology. In this sense the term 'technomethodology' is an identification of the need for the incorporation of ethnomethodologically informed accounts of the working division of labour that places the actor's perspective at the centre. The object of the ethnographic exercise is therefore to provide what Gertz (1975) terms as 'thick descriptions', which will inform the design and the implementation of the new information technologies.[6]

CSCW can inform the development and implementation of new information technologies that are directed towards teaching and learning environments in the social sciences. I do not wish to argue that the these environments are the same as commercial and industrial settings which so far have been the foci of CSCW. At a generic level university teaching and learning settings are similar insofar as they also require cooperation, coordination and collaboration to accomplish work tasks. And whilst the work carried out in university teaching and learning settings is arguably, often of a more individual nature, CSCW has attended to the issue of individualistic work.[7] If we treat the concept of 'doing work' as the active process of 'sense-making' that individuals undertake in settings, then the same kinds of issues that impinge upon actors in commercial work settings are also present in university teaching and learning settings.

Computer Supported Cooperative Work and Teaching in the Social Sciences
It is certainly the case in Britain that the desire to incorporate technology into higher education teaching and learning has been firmly placed on the higher education agenda. This situation is both desirable and essential to the future development of university education. If in the social sciences we wish to move from using computers as tools, to a scenario where we develop computer based learning

technologies I believe that it is fundamental to incorporate non-technical knowledge. This will lead to a more circumspect and strategic development of information technology than could be achieved by enthusiasts or technical specialists.

The development of Internet based teaching and learning resources for the social sciences must proceed from clear sets of pedagogical requirements. This is not to argue that in any sense objective sets of criteria that form typologies for effective resources exist. The resource requirements in various settings will be context specific. The role of a CSCW strategy, that is technomethodologically informed, is an attempt to uncover the pedagogical requirements of the Internet computer based resource that is being developed. This is the level of sophistication that is required to develop high quality Internet based teaching and learning resources that will be useful and used in social science departments.

A CSCW approach to the design and implementation of Internet based resources for social science teaching and learning will be liberating. The need for a clear understanding of teaching and learning environments is critical. Technomethodologically informed CSCW, when brought to bear upon the design and implementation of Internet based resources for social science teaching is an attempt to improve what Gurdin (1988) and Bannon & Harper (1991) term as the 'distinctly random success' of new information technologies.

Bibliography

Anderson, R., Hughes J. A. & Sharock, W.W. 1989, *Working for Profit: The Social Organisation of Calculation in an Entrepreneurial Firm* , Gower Aldershot

Bannon, L. 1989, *'From Human Factors to Human Actores: The Role of Psychology and HCI Studies in Systems Design'*, Computing Services University College Dublin.

Bannon, L. & Bodker, S. 1989, *'Beyond the Interface: Encountering Artifacts in Use'*, Department of Computer Science Aarhus University.

Bannon, L. & Schmidt, K. 1991, *'CSCW: Four Characters in Search of a Context', Studies in CSCW: Theory, Practice and Design*, in J Bowers and S Benford, Elsevier Amsterdam.

Barnard, P. & Gurdin, J. 1988, *'Command Names', Handbook of Human Computer Interaction,* in Helander, M. (ed), North Holland Press Amsterdam.

Beynon, H. 1984, *Working for Ford*, Penguin London.

Blackler, F. & Osborne, D. 1987, 'Designs For The Future:

I.T. & People', *British Psychology Society Conference Leicester.*

Brooks, R. 1990, 'The Contribution of practitioner case studies to human-computer interaction science', *Interaction With Computers*, Vol. II No.1 April.

Button, G. & Dourish, P. 1996 'Technomethodology: Paradoxes and Possibilities', Technical Report EPC-1996-101, Rank Xerox, Cambridge.

Carroll, J. 1987, *Interfacing Thought: Cognitive Aspects of Human Computer Interaction*, MIT Press Cambridge.

Carroll, J. 1989, 'Evaluation, Description and Invention: Paradigms for Human Computer Interaction', *Advances in Computers*, Vol. 29.

Clark, B.M. & O'Donnell, S. 1991, 'Computer supported co-operative work', *British Telecom Technical Journal*, Vol 9 No.1 January pg. 47-56.

Daniel, J. 1993, 'The Challenge of Mass Higher Education', *Studies in Higher Education,* Vol 18(2) pp.197-203.

Garfinkel, H. 1956, 'Some sociological concepts and methods for psychiatrists', *Psychiatric Papers,* Vol 6 pp.181-195.

Garfinkel, H. 1967, *Studies in Ethnomethodology*, Prentice Hall New Jersey.

Garfinkel , H. 1986, *Ethnomethodological Studies of Work*.Routledge and Kegan Paul London.

Garfinkel, H. 1991, 'Respecification' , in Button, G. *Ethnomethodology and the Human Sciences*, Cambridge University Press Cambridge.

Gayle, V.J. 1991, 'CSCW and Individual Work: An investigation of work distribution in a London taxi firm', Unpublished Masters Thesis, Department of Sociology University of Lancaster.

Geertz, C. 1975, *'Thick Description' In the Interpretation of Culture*, Hutchinson London.

Gray, W. & Atwood, M. 1988, 'Review of Interfacing Thought', *Interfacing Thought: Cognitive Aspects of Human Computer Interaction*, in Carroll, J. *Interfacing Thought: Cognitive Aspects of Human Computer Interaction*, MIT Press Cambridge.

Grief, I. 1988, "Remarks in panel discussion on 'CSCW: What does it mean?' *CSCW '88 Proceedings of the conference on CSCW*, Sept 26-29 Portland Oregon.

Hammersley, M. & Atkinson, P. 1983, *Ethnography: Principals in Practice*, Tavistock London.

Harper, R. & Randall, D. 1992, 'Rogues in the Air: An Ethnomethodology of 'Conflict' in Socially Organised Airspace', Technical Report EPC-1992-109, Rank Xerox, Cambridge.

Harper, R., Hughes, J., Randall, D., Shapiro, D., Sharrock, W. (forthcoming), *Ordering The Skies*, Routledge.

Hirscheim, R, & Klein, K. 1989, 'Four Paradigms of Information Systems Development', *Social Aspects of Computing*, Vol 32 No. 10 October pg. 1199-1216.

Hobbes, D. 1988, Doing the Business: *Entrepreneurship the Working Class and Detectives in the East End of London*, Clarendon Oxford.

Luff, P. & Heath C. 1990, *Collaboration and Control: The Introduction of Multimedia Technology on London Underground*, Rank Xerox.

Lyytinen, K. 1989, 'Information Systems Failures: A survey and classification of empirical litterature', *Oxford Surveys in Information Technology*, Vol.4 pg.257-309.

Pollert, A. 1981, *Girls, wives, factory lives*, Macmillan, London

Suchman, L. 1987, *Plans and Situated Actions: The problem of human machine communication*, Cambridge University Press Cambridge.

Thimbleby, H., Anderson, S. & Whitten, I. 1990, 'Reflexsive CSCW: Supporting long-term personal work', *Interacting With Computers*, Vol.2 No.3 pg. 330-336.

Willis, P. 1987, *Profane Culture*, Routledge London.

[1] It is important to note that there is no clear or coherent answer to the question 'what is, or was, the goal of HCI?'. One of the more traditional answers to this question is that HCI intends to provide methods and matrices for evaluating the usability of computer systems. This stems from what can be loosely termed the 'human factors' approach. This is in contrast to cognitive scientists who argue that HCI is a work bench for the application of cognitive psychology to a real problem domain. Computer scientists assert that HCI must help to guide the definition, invention and introduction of new computing tools and environments. This argument is to some extent a product of the exigencies of the computer industry. The point here is to illustrate that HCI is a diverse discipline with fragmented foci and interests, a feature not often drawn out in HCI literature.

[2] For Example Beynon (1984) examined working at Ford, Hobbs (1987) studied detective work in East London and Pollert (1981) treated the working lives of women in a factory setting.

[3] Garfinkel states that 'the policy is recommended that any social setting be viewed as self-organizing with respect to the intelligible character of its own appearances as either representations of or as evidences-of-a-social-order. Any setting organizes its activities to make its properties as an organized environment of practical activities detectable, countable, recordable, reportable, tell-a-story-aboutable, analyzable - in short, *accountable*' (Garfinkel 1967 p.33).

[4] See especially Garfinkel (1986 & 1991).

[5] The conception of the egological organisation is motivated by the desire to provide descriptions and analysis, but raises a deep methodological question. Sociological descriptions like other theoretical accounts are thematically constructed. The methodological question at issue in this instance is that employing this approach is an attempt to provide a third person account of first person experience. This does not mean incorporating first person accounts into sociological depictions, rather a sociological re-constitution of that experience is required. The concern is not with particular people's experience, but with the organisation of experience, as it is encountered in social life, as a readily accountable, known and shared schemes of interpretation (Anderson, Hughes and Sharrock 1989).

[6] An example of such an attempt is Harper et al (forthcoming) which is an account of technology and air traffic control as part of an inter-disciplinary attempt to design and implement a technological system.

[7] My own work on London Taxi drivers (Gayle 1991) and the work of Thimbleby et al (1990) are two examples.

* Paper presented at the IASSIST/IFDO 1997 Conference, May 6th May 9th Odense, Denmark.

Vernon J. Gayle, Department of Applied Social Science, University of Stirling, Scotland. Email: vg1@stirling.ac.uk

Supporting Data Users in a World of Changing Technology

by Margaret Ward*

Changing technology has created many challenges for today's data suppliers and this paper will begin with a brief introduction regarding recent technological changes. It will then go on to look at the role of the 'User Services' section in the UK Data Archive and how this has now been split to create a new User Support role before finally raising some questions as to the direction user services will be taking in the future.

Introduction
We are all aware that technology has advanced rapidly over the last few years with a huge growth in dispersed computing systems, such as personal computers, and the use of the associated networking facilities especially the World Wide Web. Along with this has come a much wider range of software for use by PCs and changes in media for data delivery with CD-ROMS in particular becoming popular.

The following graph shows the media used for the delivery of data from the UK Data Archive between 1993 and 1997 and how it has changed over this period of time.

The UK Data Archive has seen the number of orders supplied increase significantly over the years. This has brought with it an increase in the number of enquiries dealt with by the User Services section. There has also been a disproportionately higher increase in queries from less experienced users, who require much more help and guidance than other types of users.
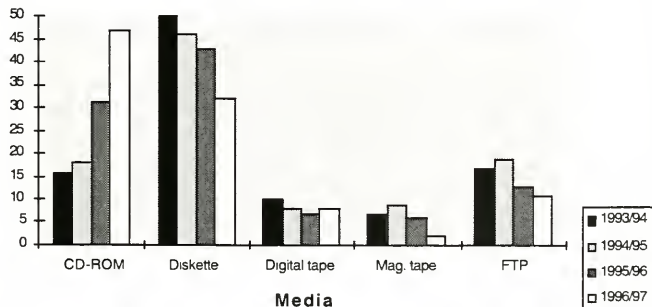
In order to provide an efficient service to users within the Archive's constrained resources it was decided to divide User Services into 'pre' and 'post' data delivery. Thus User Services staff now specialise in assisting users before they order data and offering support to users who have queries after their data and documentation have been delivered.

Pre-order queries in the UK Data Archive
Two User Services staff focus upon enquiries before any data or documentation have been delivered. These tend to fall into the following categories of 'pre-order' enquiries:

- General information
- Ordering data
- The forms required and how to complete them
- The datasets held by the Archive
- The costs which may be involved
- The formats available
  - The media available

These sort of enquiries have been received by the Archive regardless of the technological changes that have taken place, but the detailed nature of these enquiries has changed as technology has changed.

Although there has always been a choice of media on which to receive data this choice has now expanded and users need advice on which would be most suitable for the dataset they are thinking of ordering. They may also ask about the formats in which a dataset can be delivered. Can it be converted to a format the user is familiar

Media used for data delivery 1993 to

Legend: 1993/94, 1994/95, 1995/96, 1996/97

X-axis (Media): CD-ROM, Diskette, Digital tape, Mag. tape, FTP

with? Will it be easily read by their favourite spreadsheet package?

These types of questions mean that User Services staff need to be knowledgeable about the various media that are available and the limitations of each. They also need to know more about the various formats appropriate for different data and whether the dataset can be converted to this format for the user. This means that it is important for User Services staff to keep up to date with changes in technology and that they should have an understanding of the various formats and media which are currently available.

Post-order' queries in the UK Data Archive.
'Post-order' queries tend to be more broader than other types of queries and often need some investigation to resolve. Therefore in order to provide a better 'post-order' service to the users of the Data Archive it was decided that one person should have responsibility for handling this type of query and take on a more supportive role. One of the advantages to users of this change is that they would have one key person to contact. This person has responsibility for ensuring that all queries received are dealt with directly or by redirecting them to the most appropriate person within the Archive.

If a query can not be answered by archive staff, and the data depositor needs to be contacted, then the post-order support person is also responsible for contacting the relevant depositor and also for keeping track of the progress of the query with them. The user is kept informed at all times of the progress of their particular query.

Examples of 'post-order' queries?

■ "I'm having problems importing some export files from the CD you recently sent me. What should I do?"
■ "I seem to have more categories for one of the variables in my file than there are labels. Could you tell me what this one means?"
■ "Could you give me some more information about this variable, I'm not sure exactly what is included in it?"

The Queries Database
To enable queries to be tracked through the Archive and in order to ensure that we improve our service in future by learning from the queries we receive, it was decided that all of them should be logged into a database. As all the queries are channelled through one person, this person has responsibility for ensuring that all the necessary details are entered before being assigned a query number. This person is also responsible for initially examining every query and dealing with it when possible or deciding who is the most appropriate person to pass it to.

The database is accessible by all members of staff. This is to provide everyone with the ability to check the current status of any query and to enable them to add any relevant information they may have as to the current status of a particular query. This is particularly important if a query has been re-assigned as any relevant information must be added to the query so that anyone can find out the current status of it.

It was decided that a Microsoft 'Access' database would be used to record the information for each query. This was decided in part because other databases within the Archive were also to be written in Access, including the new order tracking system, and it would therefore be easier to link in any common information such as names and addresses. But what information should be recorded? Discussions took place with different members of Archive staff as to the use which would be made of the information recorded in order to determine the choice of fields.

The information recorded for each query is as follows:

■ Date query is logged onto the database
■ Name of person logging the query
■ Name and address of person reporting problem
■ Priority of query (assigned by the archive on a High, Medium or Low basis)
■ Study number
■ Order number
■ Brief details of the problem
■ Name of person to whom the query has been assigned

As more information is gathered regarding each query further comment fields can be added, which are also dated, thus allowing the progress of a query to be seen at any time. When a query has been resolved details of the resolution are entered. A brief note of any action that has been taken is also recorded, examples of which are:

■ Referred to depositor
■ Advice given
■ Re-order entered
■ Data re-acquired

A field containing a category relating to the type of query will be added shortly. This was not included when the database was designed as we wished first to monitor the types of queries we received. We have used this information in order to decide on a standard list of query types. A standard list will be used for analysis purposes in order to identify the types of problems we receive.

The queries database has been operating for the past seven months and the following graph shows the number of queries logged per month.

Advantages of, and information provided by, the queries database

Logging 'post-order' queries enables problems to be tracked, provides the Archive with information which can be used to enhance and improve the support given to its users, and also results in statistics on the performance of the Archive as detailed below:

■ No query is lost. As all queries are logged in one database none can be lost within the Archive as could happen if the query details are not recorded and passed on orally, or emailed, from person to person. The information is also being maintained in a database accessible to all staff.

■ Depending on the resolution to a query additional information may be needed to enhance the documentation for future users of that particular dataset. An example of this would be where a user has identified a variable with no variable or value labels. Once information has been received from the depositor this is added to the documentation supplied with that particular dataset so that other users do not experience the same problem.

■ The database also provides information on Frequently Encountered Problems (FEPs) as opposed to Frequently Asked Questions (FAQs)! By logging all queries it is possible to identify the types of problem users are having and enables us to make appropriate improvements to the Archive's services in order to reduce future problems. For example if a number of users are having difficulties reading files from CD-ROMs then perhaps there is a problem with the way they are being written, or the way a users' hardware or software handles CDs. Problems with extracting compressed files from floppy disks may be due to users not understanding the instructions they have been given, or possibly to inadequate instructions

provided by the Archive. These types of problems are investigated and action taken if it is thought necessary. The database allows a range of problems to be identified, monitored and resolved.

■ Particular datasets which result in a number of queries can also be identified. These could be examined further to see if there is anything 'peculiar' to these datasets. Perhaps they are only provided in a certain format and this is proving to be problematic for users. Alternatively these could be known to be difficult to use so perhaps additional documentation is needed to make them more 'user-friendly'.

■ Queries which have had to be referred to depositors can also be identified. This allows the Archive to identify which datasets have needed extra information from the depositors before they can be used more easily by the Archive's users. Hopefully this can assist us to improve the acquisition process for data in the future.
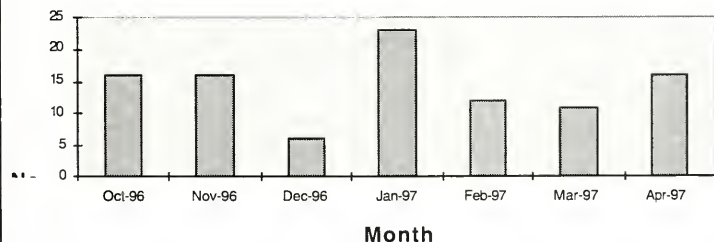
■ The Archive can also identify which of its users repeatedly report queries! Perhaps some users need more support than we can reasonably provide and we might involve local organisational representatives to assist them.

Using the information provided by the database
As mentioned above, a vast amount of information can be gleaned from the queries database. But how can the Archive use the information obtained effectively?

Creating additional notes to be added to existing documentation has already been mentioned, however perhaps this information should be made more widely available. Perhaps we could utilise the Archive's Web pages more effectively to distribute information. These would be accessible by everyone and could be promoted as a place to look to first before contacting the Archive. This may also be particularly useful for datasets used by a large number of users in several countries; the IMF databank, supplied by ICPSR for example. These pages would have to be organised in such a way as to enable users to find the information which was relevant to them quickly and easily, but would be effective in providing information to a large

Number of queries logged in database

Month

number of people.

The Data Archive could link additional useful information to the BIRON (our on-line catalogue) entry for a dataset. Users who are using a particular dataset could then look at the relevant part of BIRON and see if there is any new information relating to that study. However it is important that any information relating to a dataset is also included with the main documentation supplied with the data. Users should not have to have to check different web pages for vital information necessary for analysing the data which have been supplied to them. It should also not be forgotten that some users do not have access to the World Wide Web, but are entitled to the same support as those that do!

The future?
What type of service will data suppliers be offering in the future? Will all data formats and media be available for users or will availability be limited? Just how much support should be given once the data has been delivered? What is a reasonable amount of time to spend on one query? Should this time be limited and should a charge be made for the help given? These are questions which I think will become important in the future.

Should this support be monitored? It has been argued that it takes longer to log a query than to answer it! In some cases this is true **but** I believe the information which can be gained by monitoring queries received far outweighs the time taken to record the details. So much can be discovered about users and the problems they have which can be used to make sure that we, the data supplier, provide them with a good support service and so ensure that as a data supplier we do indeed have a future!

* Paper presented at the IASSIST/IFDO 1997 Conference, May 6th May 9th Odense, Denmark.

Data Services Assists Teaching and Research: Delivery of Data Services via the World Wide Web

by Brian Kroeker

Overview

So far in it's history, the Data Services Division at UBC has played a passive role in the teaching and research that goes on at UBC. Patrons have come to us as the machine-readable data experts for many types of files in the different subject areas covered by the teaching and research efforts at UBC. There has been a lot of personal interaction with patrons as they struggle to learn various computer systems and access retrieval methods to get the data they want. With increased improvements in computer networking technologies that have taken place in the last few years, we are seeing patrons in a new light as their needs change. The UBC Data Services Division is thus becoming more proactive in the teaching and research efforts of our patron groups. We are changing our role in response to shifting patron needs; to offer services that will assist patrons towards self-sufficiency at information discovery and retrieval rather than being dependent on personal interaction with a data librarian. Such a move requires close contact with faculty and students to develop the tools that they need to achieve their goals. It also makes us a more active participant in the learning proces that goes on in the classroom.

This paper will discuss some of the issues involved in providing access to data files over a computer network. It will be shown that there are different, identifiable user needs to searching out and aquiring data, and that we can employ mechanisms that can help us handle some of these needs. Specific examples of changes that have been implemented and are envisioned for the future at UBC will be examined.

Reasons for Change

There are two common themes in many Canadian public institutions: an external demand for increased service from patrons coupled with an internal desire from staff not to allow currently perceived levels of service decline. During the first three decades following World War II, there was a great desire on the part of governments and their citizens to grow, to build, to do new and innovative things. As a result, governments around the world went deep into debt, spending on and lending money for capital projects to build infrastructure like roads and schools and industrial systems, things that were believed to be needed to drive our economies. Governments and the public alike became addicted to public debt, which left citizens with a society-wide attitude that they had only to ask for something, and the government would provide. As universities in Canada are mainly funded by the government, the idea of perpetual growth in services has been reflected at university libraries as well. The situations which fueled some of the world-wide desire for growth have since shifted, leaving governments everywhere in the position of no longer being able to justify the current levels of public debt. Many members of the public though seem to be left with a desire for continued 'upward' growth and the constant provision of more services for which the existing funds are no longer able to provide.

Many institutions reacted initially to the trend of downward levels of funding by introducing new ideas about the use of technology in the workplace as well as methods of reorganizing operations. This was done in the hopes that greater efficiencies would lead to the maintainance of current levels of existing services, or even create opportunities for continued growth. The promise that technology and organizational restructuring by themselves would lead to continued growth has in large part failed to lead to desired improvements in overall levels of service in many academic libraries. Some of the tools that were deployed towards that goal however, have lead to new services and ways of providing service. Important lessons that have been learned in the last few years from the organizational standpoint are:

- Know your patrons/clients/customers and what their needs are. Stay in touch.
- Know your available and potential resource limitations. Letting your patrons know as well helps them to understand your position better.
- Know what other institutions are doing.

We know now that continued growth of services will not happen in the current climate of downward or static funding levels. Less money and fewer staff means lower levels of service. Since patrons and service providers alike wish improved service, some improvements will likely be made in some areas, but the trade-off is that some existing services will be reduced or discontinued.

The last point to realize why institutions must change is that patron/client/customer needs change. Many public institutions have operated and continue to operate on the principle that they provide a service to the public which does not change unless the institution itself changes. This attitude ignores the importance of the patron's/client's/customer's evolving needs in providing service. At UBC as well as elsewhere, technology and new teaching methods are having an impact on how our patrons both currently use, and desire to use our services. We need to understand how patron needs are changing, so that we may change with them.

The remainder of this paper will outline what the staff of UBC Data Services are doing to bring about improvements in service for their patrons.

Identification of Patrons and Their Needs - Are We Being Effective?

Data Libraries in general are faced mainly with two different types of patrons, those who need to perform extensive research, usually over significant amounts of time, and those who need "just a few numbers" or facts. Faculty members and graduate students fall into the first category, along with professional researchers working for the media, law firms, and other information consuming companies. These researchers do not usually require significant amounts of staff time by the time they visit a data library as the academic way of life has taught many of them the art of self-reliance and asking detailed, direct questions. This group of patrons may need some initial training in procedures, but many of them are willing to jump through many hoops all by themselves to get what they want. If they can be pointed in the right direction, they will do the rest, so starting them off on the right footing is important.

Staff time can be consumed in large quantities by those patrons still being trained in the academic way of life, or for those who are not traditional students at all, i.e. members of the general public. Patrons who "just need a few numbers" have often been directed to an archive or library by someone more experienced such as a faculty member or former/current user of the institution themselves. This group of patrons most often encompasses undergraduate students, members of the media, and the general public. Members of this patron group may arrive on the doorstep of a data library with unrealistic expectations of service and end up consuming large amounts of staff time. This group needs to know exactly what the collections of the Data Library hold, as well as services provided. They also need to know where to go if the Data Library can't answer their needs, i.e. where is the next most likely place to go?

The UBC Data Services Division has always had a mandate to provide data services to the entire campus, but due to

staffing levels and availability of subject matter, the scope of the division's activities have always been focused on the Social Sciences rather than on every subject area UBC offers. Within the Social Sciences, printed reports in the subject divisions serve a growing number of students and researchers alike, and UBC's data-oriented library staff are too small in number to be familiar with the bulk of them, not working with them every day as others do. It is no longer possible, if indeed it ever was, for Data Services staff to effectively identify for patrons sources of information that they require or could use, especially print sources. As the growth of electronic sources of information continues, it is becoming apparent that Data Services does not have the staff to provide comprehensive knowledge of all subject areas under its mandate. Patrons need a single point of service for their subject information needs if this is possible.

The bulk of UBC's Data Services collections come from statistical bodies, both governmental and non-governmental. As well as raw, or primary research data, some of these data producing organizations also create and redistribute reams of machine-readable (static tables and reports) and print materials based on their raw data. In some cases their raw data are merged with the data of other organizations, leading to even more opportunities for analysis and reports. These printed and machine-readable report style materials are often more easily consumable information products than raw data products, which require specialized training to use effectively. The report style products are used by those patrons who wish to perform secondary analysis, i.e. to just "look up" what the statistical producers and others have decided would be of most common interest. Patrons have two needs in this sea of information products:

■ They need experts who can guide them to the right sources of information for their needs.
■ They need experts who are familiar with the various formats the information can come in and who can guide them in accessing the information.

How Has UBC Data Services Filled Patron Needs?

Within both patron groups identified in the previous section at UBC, Data Services staff members were answering many of the same questions over and over again. We wondered if we were performing an effective service here. It was a well thought of service, given the positive comments and praise we have received over the years, but was this the best way to serve our patrons? Could this time have been better spent, but still provide our patrons with what they required?

We turned to the technology of the WWW to answer these questions. We designed what we think is a fairly informative WWW site, one that answers many of those basic questions that we used to get over and over. We then

began directing faculty and students to this site when they had basic questions, instead of answering their initial queries for information. We also coordinated somewhat with faculty to make sure that this site contained the kinds of information they required for their research needs, and to pass on to their students, though this aspect requires more development. On top of this, we have educated our colleagues in other divisions about our WWW site and what it contains. The result after over a year of operation is that personal patron visits to the Data Services division are way down, but visits (hits) to our WWW site are very high, in the hundreds of hits per month range from UBC addresses alone. Feedback we have had from students and faculty indicates that people like this technology and are relying on our site to provide them with basic information. We feel that when people can read information when they need it, they will obtain a better grasp of what to do to get their data, as opposed to a data services staff member telling them what to do orally. When we didn't have the WWW site, patrons would often come into the Data Services Division time and time again for clarification and further understanding of procedures. This seems to happen a lot less frequently now.

Data Services used to be a low-profile reference point to which patrons were referred when they had already seen the alternatives of the specific subject areas. We have raised our profile over the years on campus however, primarily via our periodically printed newsletter, contacts with faculty members, and our WWW site. The downside of becoming better known is that we are finding that some patrons consider us the first point of contact for their information needs. Given the above identified problems in the section dealing with user needs and our effectiveness, maintaining a separate reference point for data services is now seen as undesireable. A more effective service for patrons should be service integration, which is discussed in the next section.

Lastly, patrons need access to data files and documentation on the campus network. UBC Data Services has made all of its holdings available on several campus unix machines through the use of the Network File System (NFS). NFS allows various machines on the campus network to share our unix machine's mountable disk. Patrons on the other unix machines around campus can then use our files as if they were locally mounted on their own unix machines. The NFS software is also available for the PC environment, but we haven't yet tried hooking our filesystem up to any PC's on campus. We haven't got the disk space for all of our files to be useable at one time in this fashion, but we can put up any file a patron wants usually in 24 hours or less by moving off or compressing other files that are not currently being used and swapping in the new file(s). A number of other institutions are using this approach as well.

This last technique allows patrons access to our data and documentation, but not user-friendly access. Since the unix machines attached to our unix machine through NFS are usually departmental machines and departments are unwilling to spend the people resources necessary to provide their students with access on these machines, undergraduate students generally have no access to our files from the departmental computers. There is a central campus machine that undergraduate students can get accounts on, but currently there is a fee associated with using it. We are using the WWW to allow all our patrons another level of access to our data files, one that bypasses the central campus machine. Details of our WWW access project is covered in following sections.

Integrating Service Points - Shifting Time To Other Needs

UBC, like many libraries, collects much printed and machine-readable information, but we have traditionally divorced their physical locations based on media types. If it was electronic, then it went to Data Services, if print, it went to the applicable subject area. This divorce between electronic and print media is becoming increasingly undesirable from the user's point of view, so we have decided that closer integration between electronic and print media in a given subject area should occur. It can be confusing for the patron who doesn't necessarily see why subject matter is separated into separate library divisions just because one set of material is in electronic form and one is in print form. The patron is more apt to want one-stop-shopping for all their subject needs, so if we can provide this convenience for the user at little cost to ourselves, then we should. By starting their search for information with the subject specialist librarian, if the patron does need to see the data specialist they should be somewhat educated in what they can expect from the institution's electronic and print collections, and what services can and cannot be provided to them. In short, data services staff will see more focused questions, and fewer of them.

A significant portion of our collection deals with electronic versions of materials already held in our Government Publications division in print form, as well as the raw data that go into the making of many reports available there, so this is a natural place for us to begin our integration efforts. There is a growing sense that the front line for library service to the general public, students, and researchers alike for all subject materials, whether electronic or print, should be shifted to the people in charge of the subject section, and away from the Data Services staff. We expect integration with other subject areas to continue as time progresses.

We have begun efforts to educate subject specialist librarians in the Government Publications Division about information products that mirror the print format, and at

least one reference librarian is being trained in microdata in conjuction with the Data Liberation Initiative, a cooperative project between Statistics Canada and research libraries across Canada. It is hoped that the people at the Government Publications Division will take over the basic reference questions from Data Services for government electronic information of all sorts, assisted by our WWW site, which will continue to grow and improve.

As already alluded to, our experience shows that a significant number of new patrons of our data services division often make the mistake of applying directly to us for information when they can be served by that which is already in print. This may be due at least in part to the popular myth that everything is now available "somewhere" in digital form, and one has only to sit down and easily gain access to the entire spectrum of human knowledge from the nearest computer terminal. It takes time to identify what subject the patron is inquiring after and to direct them to resources that might assist them. A shift away from Data Services staff to the subject specialist librarians providing initial contact information will probably get the patron directed to the resources they want, sooner. Even if the patron does need something in electronic form, they may benefit by noting what is in print form. Patrons with complex queries or patrons needing a different level of technical support will still be able to see a data services specialist.

Where To Reallocate Freed-Up Resources
What any individual institution can do for its patrons is a matter of investigation - asking a lot of questions. We have identified our users and decided that their teaching needs were not being met. We are in the process of implementing some of the changes mentioned above, in order to free up some people resources to focus on user requirements.

The following arguments and rationalizations hinge on the supposition that the World Wide Web is available to everyone at the institution, or very soon will be. The UBC campus is very well connected to the Internet and the campus is very nearly completely networked internally as well. This may not be the same state of affairs at all institutions everywhere, but we believe that UBC is close to the forefront of having everyone on campus able to access the WWW via what is currently a reasonably high-speed, high-bandwidth connection. It is the author's belief that currently less well connected institutions will be proceeding as we are proceeding in the years to come, with greater connectivity and WWW access for everyone as the cost of providing this service declines or is met by an internal reorganization of priorities.

If the Data Library is eliminated as a separate service point for reference, we will immediately free up some of the Data Services staff time as they will not be required to spend as much time with patrons as before, but the line-ups at the

main Reference Desk may get a little longer. The trade-off is that a fair bit of time savings may be realized by the Data Services staff. The question then becomes: How best then to make use of these freed up resources?

One use for the time of the former Data Services staff would be to coordinate projects within the institution to provide patrons with WWW access to electronic data files. This would require liaison with both research and teaching faculty, to see what sorts of things could be useful and what is required for themselves and their students. Many data files that data libraries deal with have either no interface programs that assist users in using them, or they rely on some kind of generalized software, like SPSS, SAS, or SHAZAM in order to extract desirable subsets of the original data files. Many of the data files housed in traditional data libraries could have surprisingly wide-ranging uses if access would become somewhat easier. Many surveys of a governmental origin, be they census data or surveys dealing with the health of the population or incidence of political opinion may be of interest to those scholars in fields traditionally outside the fields in which the surveys were taken. Non-traditional and traditional data users alike need easy access to the Data Library's data and documentation, and only through the WWW is this likely to happen, at least until something comes along to supplant the WWW.

Another proposition would be to assign the former Data Services staff to the function of WWW publication needs. This is not incompatible with the above goals of WWW access to the current data library machine-readable data files and documentation. The reasons for diverting resources to WWW publishing of library information are two-fold:

■ The WWW is becoming an important method of mass dissemination of information in electronic form.
■ The WWW is increasingly being used in teaching, for distributing information back and forth.

Many libraries would like to get resources directed towards having some sort of presence on the WWW, and the freeing up of Data Services staff from reference work should provide both badly needed computer expertise and people time in this critical area. Many libraries are facing the reality that they need a WWW presence, but that they do not have the resources to do a good job in this area. Most institutions depend on a very small number of dedicated, and often quite separated individuals to take the bull by the horns and do what they can as far as creating a WWW presence for their institution or division. What usually comes out of this approach is a patchwork solution, where each branch and division goes their own development way, some looking good, some looking poor, some updating their information frequently, and some, especially if a key individual leaves, not updating information for long periods

of time.

A group dedicated, or at least partly dedicated, to providing at least a coordinating/helping role in publishing on the WWW for the organization would go a long way towards helping out the organization. If well designed, WWW sites can be used for many types of general reference questions that patrons may have, thus reducing the waiting time in the line-up at the reference desk that may have built up temporarily when Data Services staff stopped doing reference work. Patrons can be directed to specific areas of a WWW site and be instructed to go through what they find there. This is not an argument for centralized control of all WWW publication for all divisions and branches within a library. The author believes that more of a guiding and nurturing role can be played by the WWW publications group, and that this will lead to better WWW sites and more dedicated involvement in developing the WWW sites by many more staff.

In direction of making better access to data library data and documentation, the Data Services staff at UBC have developed a prototype system for allowing access to data files and documentation in a friendly, easy to use manner, utilizing the WWW. We hired a student programmer for a summer and had him program a WWW data extraction interface to the data file and documentation for the *1993 Survey of Consumer Finances - Individuals Aged 15 Years and Over, With and Without Income* survey from Statistics Canada. This type of work is not necessarily new, others are involved in the same sort of enterprise, however we have gone to great strides to present the information in as friendly and user-obvious a manner as possible. We took the codebook as produced by Statistics Canada and marked it up in HTML. This was a fairly time consuming and tedious process, as the data vendor sent us the codebook in WordPerfect format and we had no way to use this file format directly on the WWW. There are now new and better conversion tools available to convert the popular word-processing formats directly into formats for the WWW, and this can only get easier. The newest and best tools preserve both tables and graphics from the word processor output, in HTML.

The patron sits down at any Internet accessible computer on the UBC campus with Netscape loaded onto it and selects a number of variables and cases from either the basic WWW form or the complete one. The basic form was envisioned to be used mainly by beginning students, and is used by instructors in courses introducing students to numeric survey data and it's analysis. Via the basic extraction form it is possible to only get a limited number of the most popular case selections, as defined by the teaching faculty, though full access to all variables is possible. The complete form of course provides access to all cases and variables, for researchers and others who may need access to more case selection possibilities than just those provided by the basic form.

After a patron completes the data selection forms, a CGI-program on our Data Library server writes an SPSS job file to the filesystem in a special directory. Periodically throughout the day, the SPSS program is run if there are job files waiting for it in this special directory. When the SPSS job completes, the user is notified via e-mail, and the output of their job is written to the anonymous ftp directory of the filesystem. We decided to limit patrons to a certain number of requests per day, as well as a maximum overall disk utilization per user. So far, we have received no complaints regarding these limits, though they were chosen arbitrarily.

The feedback we have had from both teaching staff and researchers has been enthusiastic and very encouraging. We intend to do more. The author is now engaged in attempting to generalize the C code used in the system, as the current code is designed around the one survey that it supports. When completed, the system should be capable of adapting new survey files and their codebooks. These should be able to be added by a non-programmer. The concepts of resource classes and objects are being used in the coding to make other extraction systems besides SPSS work in the background for data extraction. A resource is in this context an extraction system, and an object is a data file and it's associated documentation. In theory and the author hopes in practice, the patron will not be able to tell how the actual objects are stored, nor what resources are being employed to act upon them.

* Paper presented at the IASSIST/IFDO 1997 Conference, May 6th May 9th Odense, Denmark

IASSIST

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY
• • • •
ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES

The **International Association for Social Science Information Services and Technology (IASSIST)** is an international association of individuals who are engaged in the acquistion, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompases hard copy as well as machine readable data

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional and international conferences sponsored by **IASSIST**.

Membership fees are:
Regular Membership, $40.00 per calendar year.
Student Membership: $20.00 per calendar year.

Institutional subcriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subcription: $70.00 per calendar year (includes one volume of the Quarterly)

Membership form